The Prospect of STT-RAM Scaling From Readability Perspective

Yaojun Zhang, Wujie Wen, and Yiran Chen

Department of Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, PA 15261 USA

Due to its fast access time, high integration density, nonvolatility and good CMOS compatibility, Spin-transfer torque random access memory (STT-RAM) becomes one promising technology for the memory hierarchy of the next-generation computing systems. In recent years, tremendous efforts have been made to reduce the switching current of magnetic tunneling junction (MTJ) for write performance and energy improvement. However, the success of write current reduction makes the STT-RAM read stability issue more prominent during the scaling of STT-RAM: following the decrease in MTJ switching current, the read current must scale accordingly to keep the disturbance on the MTJ resistance state at a minimum level. If the MTJ resistance and/or TMR values do not increase proportionally, the effective sense margin of STT-RAM will degrade, leading to a higher sensing error rate. In this work, we quantitatively analyzed the impacts of existing MTJ scaling rule on the readability STT-RAM, including both read disturbance and sensing errors. We also presented the importance of selecting an optimal read current for maintaining the readability of STT-RAM under the current scaling trend.

Index Terms—**STT-RAM, reading disturbance, sensing error, perpendicular MTJ.**

I. Introduction

SPIN-TRANSFER torque random access memory (STT-RAM) is a promising technology for the embedded memories and on-chip cache applications. Compared to other competitors, STT-RAM offers nanosecond access time, small memory cell size, good CMOS process compatibility and scalability [11]. As shown in Fig. 1(a), STT-RAM relies on the magnetization programming of MTJ (magnetic tunneling junction). A popular one-transistor-one-MTJ (1T1J) STT-RAM cell structure is shown in Fig. 1(d), where the NMOS transistor size must be large enough to supply sufficient switching current to the MTJ.

As technology scales, tremendous efforts have been made on reducing the MTJ switching current and improving the energy and performance of STT-RAM writes. Many new MTJ devices, e.g., the MTJ with perpendicular magnetization [7], dual tunneling barrier [9] and speculating layer, have emerged. These new device structures achieved one order reduction on the switching current magnitude compared to conventional in-plane MTJs. The recent experimental results show that such improvements realized a fast STT-RAM write access time as low as 3 ns [10]. However, the MTJ device variability incurred in the manufacturing may result in the variation of the sensing margin, leading to a relatively long sensing time (>2 ns) when the sensing margin is small. However, a long sensing time may induce the read disturbance issue when the magnetization state of the MTJ is flipped during the read. As the MTJ switching current reduces, the read current and/or the sensing time must be scaled accordingly to maintain a low read disturbance rate. In the other words, the conflict between the read performance and the read stability emerges as a critical issue following the scaling of MTJ devices.

In this work, we quantitatively analyze the robustness of the STT-RAM write operations tracking down the current scaling path of MTJ devices, i.e., from 45 nm to 22 nm. Both persistent errors and nonpersistent errors, which are induced by process variations and thermal-induced MTJ switching randomness, respectively, are included in our simulations. Finally, we also present the importance of finding the optimal read current to enhance the STT-RAM read reliability at different technology nodes.

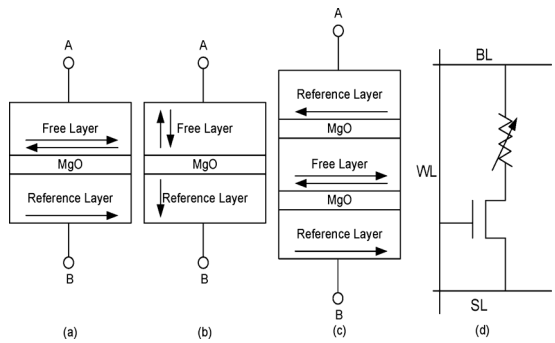

Fig. 1. (a) Conventional STT-RAM. (b) Perpendicular STT-RAM. (c) Dual Barrier Layer STT-RAM. (d) 1T1J STT-RAM cell structure.

The rest of our paper is organized as follows: Section II introduces the MTJ device structures that may be used at the scaled STT-RAM technology nodes; Section III describes the scaling trend of sensing margin and the relevant technical tradeoffs; Section IV gives our experimental results; Section V concludes our work.

II. STT-RAM and Switching Technologies

A typical STT-RAM cell design includes one MTJ and one NMOS transistor, as shown in Fig. 1(d). As the data storage device, the MTJ can be switched between two resistance states by applying a polarized write current through it. When the magnetization directions of the free layer and the reference layer are parallel or anti-parallel, the MTJ is in its low or high resistance state (say, R_L or R_H).

There are two general approaches to improve the STT-RAM write performance—increasing the MTJ driving current or decreasing the MTJ threshold switching current (for a fixed switching time). As an efficient way to raise the MTJ driving

Manuscript received March 02, 2012; revised June 01, 2012; accepted June 01, 2012. Date of current version October 19, 2012. Corresponding author: Y. Zhang (e-mail: yaz24@pitt.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMAG.2012.2203589

current, increasing the NMOS transistor size inevitably results in STT-RAM cell area overhead and degrades the integration density. In magnetism society, the research is mainly focusing on the reduction of MTJ threshold switching current.

The intrinsic threshold current density J_{C0} of conventional in-plane MTJ devices [see Fig. 1(a)], which is the minimum current required to flip the MTJ resistance in the absence of any external magnetic field at 0 K, is given by [6]

$$J_{C0} = \left(\frac{2e}{\hbar}\right)\left(\frac{\alpha}{\eta}\right)(t_F M_S)(H_k \pm H_{\rm ext} + 2\pi M_S). \quad (1)$$

Here, e is the electron charge. α is the damping constant, M_S is the saturation magnetization, t_F is the thickness of the free layer, $\hbar$ is the reduced planck's constant, H_k is the effective anisotropy field including magneto crystalline anisotropy and shape anisotropy. $H_{\rm ext}$ is the external field. The MTJ critical switching current I_{C0} is presented as: $I_{C0} = J_{C0} \cdot A_{\rm MTJ}$ where $A_{\rm MTJ}$ is the cross section area of the MTJ elliptic cylinder. I_{C0} is determined by the material property of MTJ (i.e., M_S, spin efficiency η, damping ratio α, etc.), the MTJ geometry property (i.e., $A_{\rm MTJ}$, t_F), and the working environment (i.e., working temperature). Many new types of MTJ devices have emerged for I_{C0} reduction by optimizing these parameters. Some examples are:

1) Perpendicular MTJ: Perpendicular MTJ devices [see Fig. 1(b)] include a free layer with a perpendicular anisotropy $H_{k\perp} > 4\pi M_S$ [6]. Thus, the H_k term in (1) is transformed to the effective perpendicular anisotropy as $H_k^{\perp} = H_{k\perp} - 4\pi M_S$. In this way, I_{C0} can potentially be further scaled because the $2\pi M_S$ factor from the out-of plane demagnetizing field is canceled by the perpendicular spin transfer torque. In the meanwhile, other material property parameters, especially spin transfer torque efficiency μ and damping α maintain comparable to those in in-plane MTJ [1]. Hence, compared to the current STT-RAM implemented with in-plane MTJs, the perpendicular MTJ-based STT-RAM (P-STT) can achieve lower write energy and/or better write performance, while maintaining the similar thermal stability.

2) Dual Tunneling Barrier MTJ: As the name indicates, dual tunneling barrier MTJ device [see Fig. 1(c)] is designed by adding an additional fixed magnetic layer on top of the free layer as an additional source of spin polarized electrons [4]. The two anti-paralleled fixed layers will enhance the effective spin torque magnitude on the free layer under the same switching current. In the other words, it will cost less current to flip the MTJ magnetization state, compared to the conventional MTJ structure.

III. Read Performance Following Technology Scaling

A. Switching Performance and Reading Disturbance

The invention of the emerging MTJ device structures leads to a fast switching performance down to a few nanoseconds [8]. However, the reduction of the MTJ threshold switching current also makes the magnetization state of MTJ easier to be flipped during the read operations.

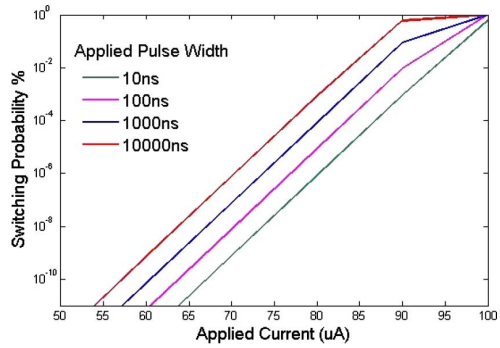

Fig. 2. Switching Probability under $I_{C0} = 100\ \mu$A.

The MTJ switching probability $P_{\rm sw}$ is a function of the MTJ critical switching current I_{C0}, the inverse of the attempt frequency τ_0, the applied current I_C and the pulse width τ_p as

$$P_{\rm sw} = 1 - \exp\{-\tau_p/\tau_0 \exp[-E/k_B T(1 - I_c/I_{c0})]\}. \quad (2)$$

We simulated the switching probability of an elliptical MTJ with the size of 45 nm × 90 nm, as shown in Fig. 2. To minimize the read disturbance during the read operations, the read current $I_{\rm read}$ must be sufficiently lower than the normal write current (which is $\sim$100 μA for a write pulse width of 10 ns). We note that considering the randomness incurred by the device parametric variations and thermal fluctuations, the actual $I_{\rm read}$ may need to set much lower than the one calculated by (2), as we shall show in the following sections. We note that material parameter fluctuations may cause the variations of the switching probability of MTJ by introducing the normally distributed localized fluctuation of magnetic anisotropy, etc.

B. Sense Margin

Following the scaling of switching current, the read current of the MTJ must decrease accordingly to suppress the read disturbance. This fact will introduce another concern on the readability of STT-RAM: In traditional current-sensing STT-RAM read scheme, for instance, a read current $I_{\rm read}$ is injected into the memory cell. The generated bit-line voltage is then compared to a reference voltage to read out the MTJ resistance state. The generated sense margin, which can be measured by the voltage difference between the bit-line voltage and the reference voltage, is proportional to $I_{\rm read} \cdot R_L \cdot {\rm TMR}$. Here R_L is the low MTJ resistance. TMR is tunnel magneto-resistance ratio. Certain sense margin must be maintained in STT-RAM read operations to overcome the device mismatch in the sense amplifier and keep the sensing errors at a minimum level.

When $I_{\rm read}$ decreases, the generated sense margin of STT-RAM may reduce if the MTJ resistance and/or TMR do not increase proportionally. The degraded sense margin may incur sensing errors if the device variation of sense amplifier is large. Since the process variations of CMOS technology become more and more severe when manufacturing technology scales, readability may replace the write failure as the limiting factor of the reliability of STT-RAM design. It is necessary to conduct a detailed analysis on the robustness degradation of the STT-RAM read operations and explore the MTJ scaling rule optimization from the readability perspective.

TABLE I
SUMMARY OF STT-RAM PARAMETERS [3]

Tech. (nm)		45	32	22
MTJ Geometry (nm)	Length	90	65	45
	Width	45	32	22
	Oxide Thickness	2.2	2.2	3.2
R_H (Ω)		2000	2500	4000
R_L (Ω)		1000	1000	1500
Transistor Size (nm)		270	150	75

IV. EXPERIMENTS AND DISCUSSIONS

A. Simulation Setup

The STT-RAM readability concern during the scaling path can be considered as the tradeoff between the read disturbance and the sensing errors: decreasing the $I_{\rm read}$ can minimize the read disturbance, however, also raise the sensing error rate. In this section, we analyze the read operations of STT-RAM under 45, 32, and 22 nm technology nodes. Based on international technology roadmap of semiconductor [5], 45 and 32 nm technology nodes will still use in-plane MTJ while perpendicular MTJ will be adopted in 22 nm technology.

Both read disturbance and sensing errors are included in our experiments: We use stochastic Landau–Lifshitz–Gilbert (LLG) equation to simulate the magnetization switching of the MTJ. Thermal fluctuation is represented by a normalized thermal agitation fluctuating field [12]. We also conduct Monte-Carlo simulations analyze the impacts of process variations on the sense amplifier's sensing margin and the sensing errors of STT-RAM. Table I shows the nominal STT-RAM cell parameters at different technology nodes, which are extracted from [3]. Following the scaling of MTJ and NMOS transistor devices, the resistance states of the MTJ increase accordingly. Predictive technology model (PTM) is used in our NMOS transistor simulation [2].

B. Reading Operation Error Analysis

As aforementioned in Section III, two major error sources of STT-RAM read operations are: 1) sensing errors; and 2) read disturbance. Here the sensing error denotes the case that the read-out voltage difference between the bit-line voltage and reference voltage is smaller than the required sense margin. In such a case, the MTJ resistance state could not be read out correctly, e.g., reading out as an "1" when storing a "0" or reading out as a "0" when storing an "1." In STT-RAM read scheme, we normally apply $I_{\rm read}$ only from one direction. Due to the asymmetry of MTJ switching, the MTJ flipping from "0" to "1" is more difficulty than the other direction. In our simulation, we assume the $I_{\rm read}$ is applied from the direction that ensures the read disturbance happening only when the cell stores a "0" (and may possibly flip to "1"). In the rest of this section, we discuss these two error sources separately.

1) Sensing Error: For a fair comparison, the same sense amplifier structure is used in the simulations under different technology nodes. The ratio between the channel widths of the NMOS and PMOS transistors are maintained the same while the channel length of the transistors are adjusted according to the technology node. We define the sense margin as the voltage difference actually generated on the two inputs of the sense amplifier. A high sensing margin generally corresponds to a low sensing error rate. Because of the increased process variations, the sense margin of the sense amplifier must increase to overcome the device mismatch in the sense amplifier as the transistor feature size reduces. Fig. 3(a) shows our simulation result of the sense margins at various technologies. When technology scales from 45 to 22 nm, the read currents (20% of I_{C0}) decrease from 34 to 15 μA. The nominal sense margins generated from the STT-RAM cell are 17, 18, and 18.8 mV, respectively. Thanks to the improvement of resistance and TMR due to the adoption of perpendicular MTJ, the sense margin even slightly increases when the technology scales.

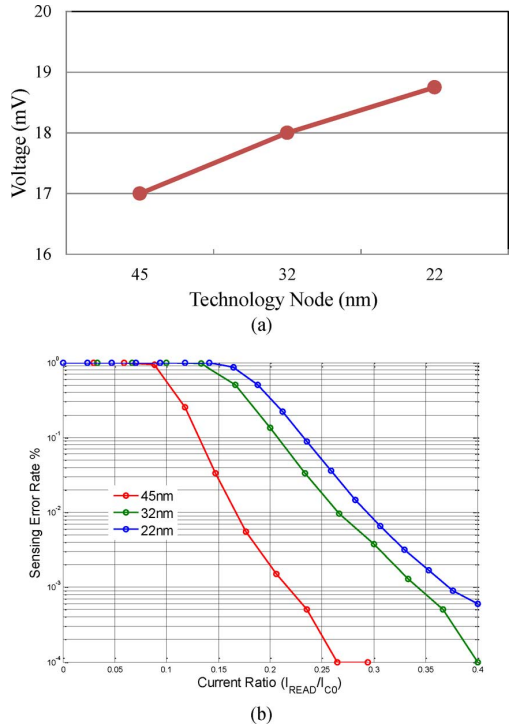

Fig. 3. (a) Sensing margin scaling trend of STT-RAM sensing circuit. (b) Sensing error rate versus reading current.

The sensing errors occur when the voltage difference on the inputs of the sense amplifier cannot overcome the device mismatch of the circuit. Fig. 3(b) shows the sensing error rates at different technology nodes when $I_{\rm read}$ changes. The device variations of both MTJ and NMOS transistor are included in our simulation. The standard deviation of geometry size is set to 5% of the nominal value while the standard deviation of the NMOS transistor threshold voltage is set to 30 mV. Following the increase of $I_{\rm read}$, the sensing error rate reduces rapidly. The highest sensing error rate always occurs at 22 nm for the same reading current ratio $I_{\rm read}/I_{C0}$, due to the combined impacts of the small critical switching current and the large variations of MTJ and CMOS devices.

We note that in our sense amplifier designs at different technologies, the transistor sizes follow a simple scaling rule, i.e.,

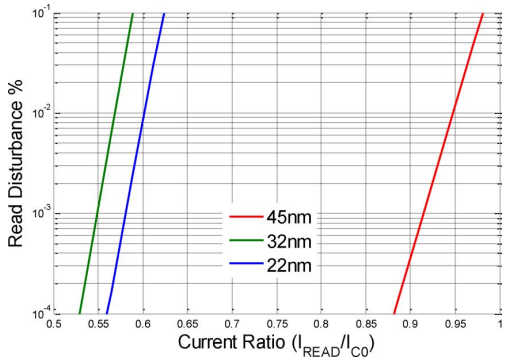

Fig. 4. Reading disturbance rate versus Read current.

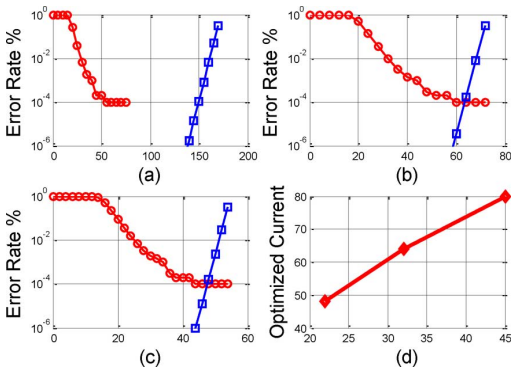

Fig. 5. Tradeoff between reading disturbance and reading current. (a) Read Current (uA): 45 nm, (b) Read Current (uA): 32 nm, (c) Read Current (uA): 22 nm, (d) Technology (nm).

multiplied by the inverse of the technology scaling factor. Increasing supply voltage ($V_{\rm dd}$) can improve the robustness of sense amplifier and reduce the sense margin. In our simulation, we set $V_{\rm dd}$ to 1.0 V for the minimization of the leakage power.

2) Reading Disturbance: Normally the read current is set to about 20% of the write currents to minimize the read disturbance. For example, the read currents of 34, 24, 15 μA are usually selected 45, 32, and 22 nm technology nodes, respectively. However, as shown in Fig. 3, such low read currents could cause severe sensing error rate in the scaled technologies. Increasing the read current, however, will raise the reading disturbance [12].

The simulated STT-RAM cell read disturbance rate under different $I_{\rm read}/{\rm I_{c0}}$ ratios are shown in Fig. 4. In all test cases, the read disturbance quickly increases when the $I_{\rm read}$ rises.

C. Read Current Optimization

Since the trends of STT-RAM read disturbance and sensing errors are opposite when the $I_{\rm read}$ changes, it is possible for us to find the optimal $I_{\rm read}$ that can achieve the minimum total read operation error rate.

Fig. 5(a), (b), and (c) shows our combined simulated results of the STT-RAM read operation error rate under different read currents at the technology nodes of 45, 32, and 22 nm, respectively. Under 45 nm technology node, there is almost no intersection between the trend curves of read disturbance and sensing errors within the concerned range. The optimal read current, which is the cross point of two error curves, is about 80 μA. When technology scales to lower nodes, both error curves shift. The cross point of two error curves increases, leading to a higher combined read operation error rate. Fig. 5(d) shows the simulated optimal point of the read current at various technology nodes.

We found that even the optimal read current is selected at 22 nm technology node, the total read error rate can be still high. Since the optimal read current is also affected by many other design parameters, i.e., the NMOS transistor and MTJ sizes, read and write pulse widths, and even sense amplifier design, all these factors may need to be optimized accordingly to achieve a lower total read error rate.

V. Conclusion

In this work, we analyzed dependency of the sensing errors and the read disturbance of STT-RAM cell on the read current under the scaled technology nodes. By taking into account the tradeoff between these two major sources of read errors, we also demonstrate the importance of read current optimization to for the enhancement of STT-RAM readability. To achieve a reasonable read error rate of STT-RAM at scaled technology nodes, many other circuit and device optimizations must be made instead of taking the current simple scaling path.

References

[1] D. Apalkov, S. Watts, A. Driskill-Smith, E. Chen, Z. Diao, and V. Nikitin, "Comparison of scaling of in-plane and perpendicular spin transfer switching technologies by micromagnetic simulation," *IEEE Trans. Magn.*, vol. 46, no. 6, pp. 2240–2243, Jun. 2010.

[2] Y. Cao *et al.*, "New paradigm of predictive MOSFET and interconnect modeling for early circuit design," in *Proc. IEEE Custom Integr. Ckt. Conf.*, 2000, pp. 201–204 [Online]. Available: http://www-device.eecs.berkeley.edu/ptm

[3] Y. Chen, X. Wang, H. Li, H. Xi, Y. Yan, and W. Zhu, "Design margin exploration of spin-transfer torque RAM (STT-RAM) in scaled technologies," *IEEE Trans. Very Large Scale Integr. VLSI) Syst.*, vol. 18, no. 12, pp. 1724–1734, Dec. 2010.

[4] G. Fuchs, I. Krivorotov, P. Braganca, N. Emley, A. Garcia, D. Ralph, and R. Buhrman, "Adjustable Spin Torque in Magnetic Tunnel Junctions With Two Fixed Layers," Tech. Rep., Mar. 2005.

[5] ITRS 2011 [Online]. Available: http://www.itrs.net/Links/2011ITRS/Home2011

[6] Y. Huai, "Spin-transfer torque MRAM (STT-MRAM): Challenges and prospects," *AAPPS Bullet.*, vol. 18, no. 6, pp. 33–40, 2008.

[7] S. Ikeda, K. Miura, H. Yamamoto, K. Mizunuma, H. Gan, M. Endo, S. Kanai, J. Hayakawa, H. Matsukura, and F. Ohno, "A perpendicularanisotropy CoFeB-MgO magnetic tunnel junction," *Nature Mater.*, vol. 9, no. 9, pp. 721–724, 2010.

[8] C. W. S. , IV, V. Mohan, A. Nigam, S. Gurumurthi, and M. R. Stan, "Relaxing non-volatility for fast and energy-efficient STT-RAM caches," in *Proc. 11th Int. Symp. High-Perform. Comput. Arch.*, 2011, pp. 50–61.

[9] N. Mojumder, C. Augustine, D. Nikonov, and K. Roy, "Effect of quantum confinement on spin transport and magnetization dynamics in dual barrier spin transfer torque magnetic tunnel junctions," *J. Appl. Phys.*, vol. 108, no. 10, pp. 104306–104306-12, Nov. 2010.

[10] A. Nigam, C. W. Smullen, IV, V. Mohan, E. Chen, S. Gurumurthi, and M. R. Stan, "Delivering on the promise of universal memory for spintransfer torque RAM (STT-RAM)," in *Proc. 17th IEEE/ACM Int. Symp. Low-Power Electron. Des., ISLPED'11*, Piscataway, NJ, 2011, pp. 121–126.

[11] A. Raychowdhury, D. Somasekhar, T. Karnik, and V. De, "Design space and scalability exploration of 1T-1STT MTJ memory arrays in the presence of variability and disturbances," in *Proc. IEEE Int. Conf. Electron Devices Meeting*, Dec. 2009, pp. 1–4.

[12] J. Z. Sun, "Spin-current interaction with a monodomain magnetic body: A model study," *Phys. Rev. B*, vol. 62, pp. 570–578, July 2000.